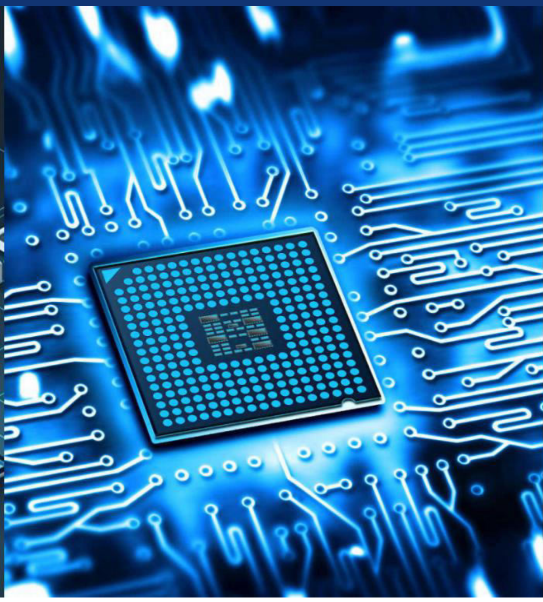
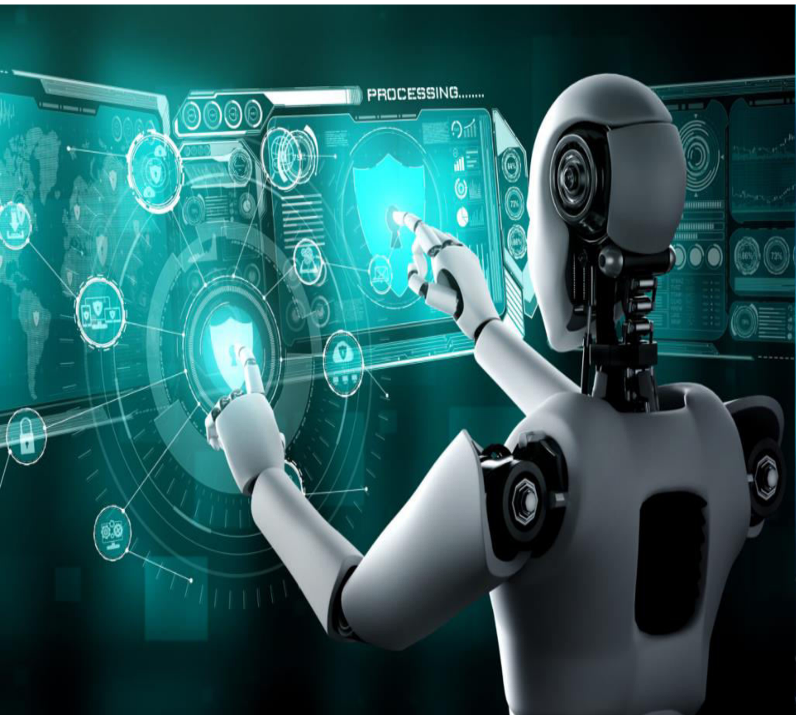
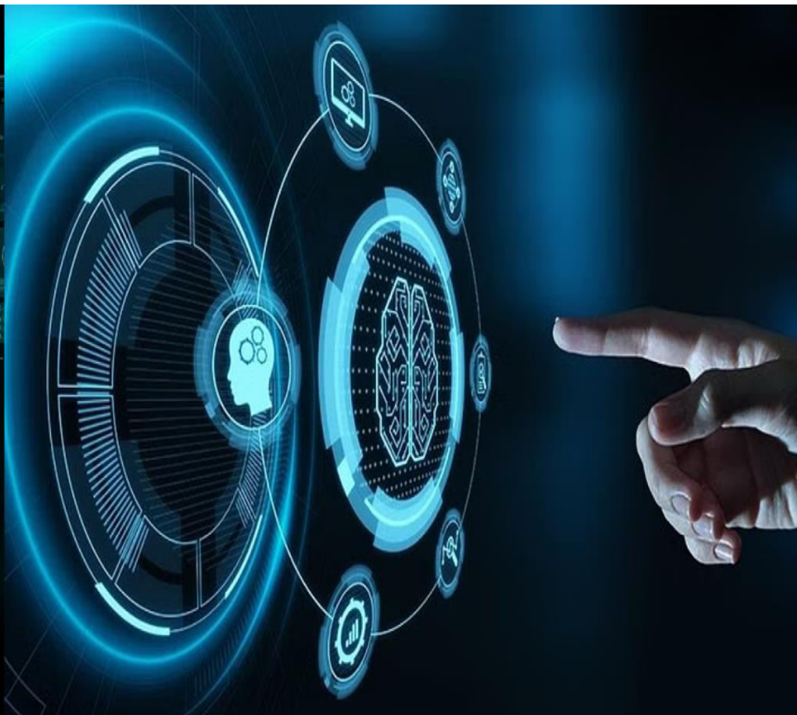
ISSN(O): 2320-9801
ISSN(P): 2320-9798


International Journal of Innovative Research in Computer and Communication Engineering

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)





Impact Factor: 8.771

Volume 13, Issue 4, April 2025

© 2025 IJIRCCE | Volume 13, Issue 4, April 2025| DOI: 10.15680/IJIRCCE.2025.1304068

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Bringing Back Old Films and Photos Back to Life

Manish Gambhirrao[1], Aditya Khapke[2], Ranjan Shitole[3], Rushi Gunjal[4], Jayshri Dhere[5], Abhilasha Bhagat[6]

B.E, Department of AI & DS, Dr. D.Y. Patil Institute of Engineering, Management and Research, SPPU, Pune, Maharashtra, India[1,2,3,4]

Dr D Y Patil Institute of Engineering, Management and Research, D. Y. Patil International University, Pune, Maharashtra, India[5,6]

ABSTRACT: An introduction of Recurrent Transformer Network is presented as a New learning-based Framework that will be able to restore films that have suffered from neglect and severe degradation during their lifespans. Indeed, classical treatment of film restoration becomes very insightful in restoring temporal coherence, generally in the case of relatively simpler artifacts such as scratches and occlusions. RTN helps by effective use of the spatio-temporal context from adjacent frames, allowing its output films to be smooth flowing and natural looking. RTN utilizes a colorization module that allows for cross illumination of textures from the keyframes throughout the length of the film.

KEYWORDS: Program Comprehension, Film Restoration, Spatio- Temporal Context, Colorization, Classic Cinema, Digital Remastering.

I. INTRODUCTION

The restoration of old films is often embroiled with complex and challenging processes, often hampered by the progressive decay of visual quality and the presence of different artifacts. Traditional approaches for restoration often cannot successfully deal with these issues, leading to terrible results that do not recreate the original essential quality of the films. It is innovative technologies like Recurrent Transformer Network (RTN) that play a part here. Using advanced machine learning techniques and a spatio-temporal context-aware framework, RTN aims to apply an enhanced restoration of film by efficiently alleviating such artifacts as scratches and occlusions and creating a coupling of coherent temporality across the shot

The advances in machine learning and computer vision have revolutionized the traditional approach to film restoration, giving rise to a level of sophistication and refinement unheard-of until now. Our series of experiments on both synthetic datasets and on real films showed clearly that RTN outperformed other existing restoration methods, providing very consistent results. It plays a very important role in the restoration of original visuals and aesthetic qualities of classic cinema, particularly to further preservation and digital remastering. The availability of the framework is expected to spur future research and exploration of this area while at the same time enabling collaboration and innovation between practitioners and researcher

Film restoration holds great importance for the preservation of cultural heritage and active enjoyment of classic cinema in time to come. Aging films are subject to physical decay, resulting in loss of such cardinal attributes as picture quality and visual entertainment. Restoration not only breathes life back into that kind of film but also retains its historical significance. Modern restoration techniques may help one to restore theatrical vibrancy and artistry in classic films, ensuring their impact on times

II. OVERVIEW OF DOMAIN

Film restoration is a vital process dedicated to preserving and revitalizing aging or damaged motion pictures, ensuring that classic films remain accessible and maintain their original visual and audio integrity. Traditional restoration methods involve meticulous manual repair of physical film reels, cleaning, and color correction to address issues like scratches, fading, and other forms of degradation. In recent years, advancements in digital technology have introduced sophisticated techniques, such as the use of Recurrent Transformer Networks (RTN). These machine learning-based frameworks

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

leverage information from adjacent frames to restore degraded films more effectively, enhancing temporal coherence and visual quality. These innovations play a crucial role in cultural preservation, allowing contemporary audiences to experience classic cinema as originally intended.

III. LITERATURE REVIEW

The literature review explores recent advancements in film restoration using deep learning techniques. Traditional methods rely on manual efforts, while modern approaches use CNNs, RNNs, and Transformer-based models for enhanced accuracy. Recurrent Transformer Networks (RTNs) have shown superior performance by leveraging spatio-temporal features across frames. Studies highlight RTNs' effectiveness in reducing noise, restoring color, and maintaining frame consistency.

These models excel in understanding both the visual content and motion continuity, which is crucial for restoring old, degraded films. Several research works also emphasize the use of GANs for realistic colorization and detail enhancement. Overall, the literature supports RTNs as a promising solution for automated, high-fidelity film restoration.

TABLE I. COMPARATIVE ANALYSIS OF KEY REFERENCES IN NEURAL NETWORKS AND TRANSFORMER ATTENTION NETWORKS

Reference	Focus Area	Methodology	Key Finding	Publication Year	Challenges	Dataset
[1]	Progressive Training of A Two- Stage Framework for Video Restoration	Two- stage framework with multi frame recurrent network and single- frame transformer.	Challenges in modeling session dependencies	2023.	Lack of comprehensive evaluations noted	LDV dataset, Custom YouTube dataset (870 4K sequences)
[2]	Image Super-Resolution Using Deep Learning	Survey of CNNs, GANs,and transformers for image super-resolution	Improved user and news representation	2023.	Relies on specific dataset characteristics	CelebA-HQ dataset.
[3]	Deep- based Film Grain Removal and Synthesis	U-Net with residual blocks and conditional GAN	Improved recommendation accuracy achieved	2022	Scalability and complexity issues	CBSD68, ‌lak24, McMaster, Set12.
[4]	Diffusion Models Beat GANs on Image Synthesis	Diffusion models using variational inference	Improved recommendation efficacy demonstrated.	2022.	Offline summarization may miss nuances	DIV2K, Flickr1024.
[5]	Bringing Old Films Back to Life	Combined restoration technique with colorization models	Enhances recommendation accuracy significantly.	2022.	Requires rich knowledge graph data.	User click history on news articles
[6]	Personalized Stylization with Pre Trained Diffusion	Stable Diffusion with additional stylization modules	Graphs enhance recommendation accuracy	2021	Challenges with data sparsity	Real- ESRGAN, SwinIR .

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

[7]	GAN-Based Deep Learning Approach for Video Super-Resolution	GAN architecture with spatial temporal convolutions	Integration of social networks beneficial	2021	Challenges in IoT-related graph learning.	ImageNet, CelebA, LSUN
[8]	Time- Travel Photography	StyleGA N2-based projection into modern image space	Improved recommendation accuracy with context	2022	Potential overfitting; needs extensive tuning	CelebA, ImageNet.
[9]	Denoisin g Diffusion Null Space Model (DDNM)	Diffusion based model using nullspace and range- space decomposition.	User modeling and content understanding enhance recommendation accuracy	2020	Difficulties in generalizing results; more researc h needed for user and content representation	MIND includes 15 million logs from 1 million users
[10]	Face Identity Disentanglem ent with Generative Adversarial Networks	GAN-based encoder- decoder architecture with adversarial training.	Outperforms traditional recommendation technique	2022.	Requires significant computational resources	Real-world user-item interaction datasets

A. Discussion/Analysis

The paper introduces a Recurrent Transformer Network (RTN) as a learning-based framework designed for restoring heavily degraded old films1 .... Unlike methods that process each frame independently, RTN leverages the information present in adjacent frames to aid the restoration process1 . This approach is based on the idea that neighboring frames contain "hidden knowledge" about occlusions and other artifacts, which can be beneficial for restoring challenging degradations in individual frames . Furthermore, the RTN architecture is designed to ensure temporal coherency in the restored video by explicitly maintaining and propagating hidden representations across frames using a recurring mechanism. This temporal modeling helps to reduce film flickering and ensures that the restoration of consecutive frames is consistent over longer periods

The literature highlights several critical research gaps in the field of film restoration, particularly in the context of employing advanced methodologies like the Recurrent Transformer Network (RTN). While RTN effectively utilizes spatio-temporal context from adjacent frames to restore damaged films, it still encounters difficulties when dealing with heavily degraded footage or complex visual artifacts. This suggests a pressing need for more sophisticated restoration techniques that can capture subtle details and nuances that traditional methods might overlook. Future research could focus on integrating hybrid approaches that combine machine learning with established restoration techniques, potentially enhancing both the visual fidelity and the authenticity of restored films.

Another significant issue is the limited availability of diverse training datasets, which can restrict the performance of restoration models. Many current models are trained on specific types of films or damage patterns, resulting in a lack of generalization to a broader range of restoration challenges. This limitation calls for the development of more comprehensive and varied datasets that encompass different genres, eras, and degradation types. Implementing data augmentation techniques or creating synthetic datasets could help bolster the model's robustness and adaptability, ultimately leading to better restoration outcomes.

Furthermore, the existing evaluation metrics often emphasize technical factors such as resolution enhancement and noise reduction, rather than the overall viewing experience or cultural impact of restored films. There is a critical need for evaluation methods that assess restoration quality from the audience's perspective, considering elements such as emotional engagement and historical authenticity. Future work should prioritize the creation of metrics that reflect the artistic and cultural significance of films, enabling a more holistic evaluation of restoration efforts.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

The RTN framework also features a mechanism for unsupervised scratch position inference . By contrasting the representation of the current frame with the hidden knowledge aggregated from adjacent frames, the network can infer the location of structured defects like scratches

Areas showing a larger discrepancy between the current.
Frame's representation and the hidden state are likely to be degraded . This approach eliminates the need for a separate defect segmentation network or manually labeled defect masks, making it more generalizable to real-world old film degradations.The inferred defect information guides the restoration process by highlighting areas that require more attention.The temporal modeling in RTN is achieved through a bi-directional recurrent network.

This architecture processes the video frames in both forward and backward directions, aggregating knowledge of the scene from both past and future adjacent frames By maintaining both a forward hidden state and a backward hidden state, the network can ensure that information from the entire temporal context contributes to the restoration of each frame, which is particularly beneficial for dealing with flickering and other temporally variant degradations The final restored frame is reconstructed by combining the information from these bi-directional hidden states.

To better integrate the temporal information with the current frame's features, the RTN employs a learnable guided mask. This soft mask is learned in an unsupervised manner by a shallow network that takes the current frame's features and the warped clean state (propagated hidden state) as input . The mask acts as a blending factor, allowing the network to selectively combine the information from the temporal priors (hidden state) and the current frame.

This is motivated by the observation that contaminants might be partially transparent, and their preserved content can still provide useful information for restoration.The paper also addresses the crucial aspect of training data by proposing a video degradation model To train the RTN effectively, a realistic method for generating paired degraded and clean video data is essential . The proposed degradation model simulates various types of artifacts commonly found in old films.

B. Additional Analysis

The Recurrent Transformer Network (RTN) performs spatial restoration using Swin Transformer blocks, which are chosen over traditional CNNs due to their superior handling of non-uniform degradations and misalignments commonly found in old films. After aggregating temporal information from previous frames and the current frame, a spatial network is needed to refine the output. The Swin Transformer, with its window-based and shifted window attention mechanism, effectively captures long-range dependencies and compensates for alignment errors caused by inaccurate optical flow, resulting in more stable training and better restoration quality. The aggregated features are first downsampled using strided convolutions before being processed by these transformer blocks.

A key innovation in RTN is its ability to detect structured defects like scratches in an unsupervised manner. By contrasting the current frame’s representation with the warped hidden state from adjacent frames, the network identifies areas with high discrepancy as likely degraded regions, eliminating the need for separate defect segmentation networks or manually labeled masks. To effectively merge temporal information with current frame features, RTN uses a learnable guided mask. A shallow network generates a soft blending mask based on the current frame's features and the warped clean state.

This mask enables selective fusion of both sources, where the final representation is computed as a weighted combination: $E(x_t)\cdot M + W\cdot(1-M)$, helping the model deal with partially transparent artifacts where some background content remains useful. For training, a custom video degradation model simulates realistic old film artifacts, including scratches, dirt, blur, noise, compression, and frame-wise temporal disturbances. This creates paired clean and degraded video sequences to supervise the learning process. RTN also supports video colorization by modifying its input to the LAB color space and predicting only the chrominance channels for grayscale frames. To prevent color fading during long-term propagation, a semantic similarity search across frames is used to generate coarse color hints, which are then concatenated with grayscale inputs and refined through RTN.

© 2025 IJIRCCE | Volume 13, Issue 4, April 2025| DOI: 10.15680/IJIRCCE.2025.1304068

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

IV. PROPOSED METHOD

The Recurrent Transformer Network (RTN) restores old film frames by not only analyzing the current frame but also incorporating information from adjacent frames, both before and after. It leverages a recurrent architecture to retain and utilize temporal information, which helps in identifying and reconstructing damaged areas such as scratches or missing content—especially since such defects may shift or vary across frames. This ensures temporal consistency and reduces flickering in the restored video. For spatial restoration within individual frames, RTN uses Swin Transformer blocks instead of traditional CNNs, as they are more effective in capturing long-range dependencies and handling spatially inconsistent degradation, even when frame alignment is imperfect due to motion estimation errors. Additionally, RTN can automatically detect the locations of scratches and other defects by comparing the current frame with the aggregated features from neighboring frames, without the need for manually labeled .

To blend the current frame's features with temporal information, RTN employs a learnable soft mask that adaptively decides how much to rely on the current frame versus the temporal priors—particularly useful when the damage is only partially opaque. To train the model, the researchers designed a synthetic degradation pipeline that introduces realistic old-film artifacts, including scratches, dust, blur, color distortion, and compression effects, into clean modern videos, thereby creating paired training data. Moreover, the RTN framework is versatile and can also be adapted for video colorization by predicting chrominance channels from grayscale frames. It propagates color information from a few reference color frames across the video using semantic similarity, enabling coherent and vivid color restoration.

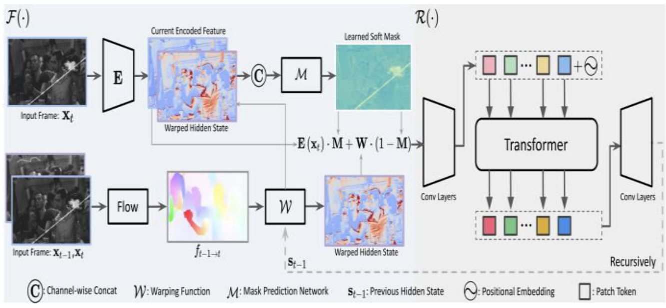


Fig. 1. System Architecture of RTN Model for Processing

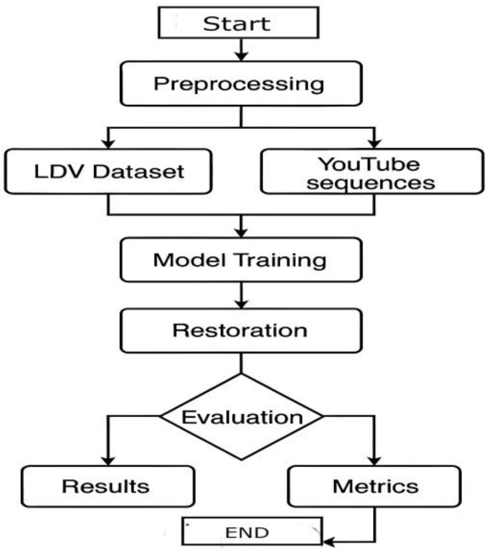


Fig 2 : System Flow Diagram for the Process

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

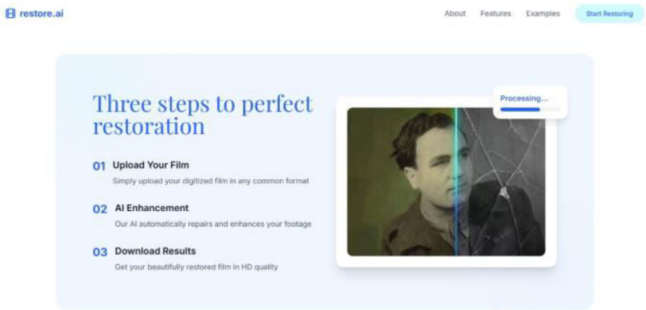


Fig. 3. Application UI Layout

A. Dataset

1. To train the RTN for film restoration, the researchers used the REDS dataset and applied their custom video degradation model on-the-fly to generate paired degraded and clean samples. They optimized the model by randomly cropping 256 patches from the dataset. For evaluation, they created a synthetic test set by applying degradations to clean frames from the DAVIS dataset, which offers diverse scenes and large camera motions. Additionally, they tested the model on 63 real old films collected online, using non-reference quality metrics like NIQE and BRISQUE due to the lack of ground truth. For video colorization, they used grayscale versions of REDS videos, predicting the colors of the first 50 frames using the 100th frame as a reference

B. Implementation

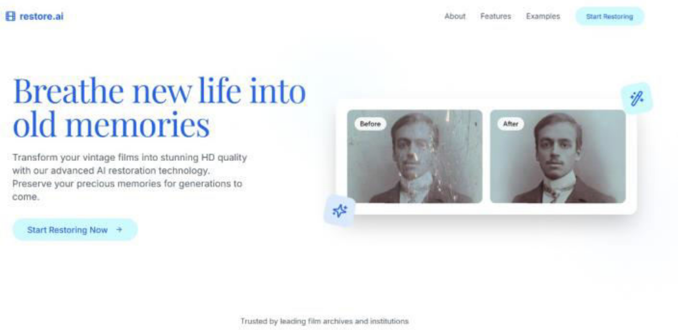


Fig. 4. The contrast of processed Image

V. LIMITATIONS

The Recurrent Transformer Network (RTN), while effective in old film restoration and colorization, faces theoretical limitations. One major challenge is the ambiguity between actual scene content and degradation artifacts. For instance, the model may confuse black scratches with scene elements like smoke, leading to incorrect restoration. This is because the network learns to recognize patterns from training data, and visually similar artifacts can be misclassified as legitimate content, highlighting the need for stronger semantic understanding.

Another limitation involves the use of GANs, which, although helpful for enhancing realism, can generate inaccurate high-frequency details or unwanted artifacts. GANs focus on producing outputs that look real rather than being exact reconstructions, which may result in plausible but incorrect restorations. Additionally, when frames are severely degraded, the network struggles to recover lost information, even with temporal context from adjacent frames. In such cases, meaningful restoration becomes theoretically infeasible due to the absence of sufficient original data.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

VI. ANALYSIS AND DISCUSSION

The paper *"Bringing Old Films Back to Life"* proposes a novel Recurrent Transformer Network (RTN) to restore old, degraded films by addressing both spatial and temporal degradations like scratches, blurs, and noise. Unlike earlier methods that process frames independently, RTN combines a bi-directional RNN for temporal context and Swin Transformer blocks for spatial restoration. It uses optical flow to align frames and a learnable mask to blend current and adjacent frame features, helping the model focus on damaged regions without manual annotations. Experimental results show RTN outperforms existing methods in both synthetic and real-world datasets using metrics like PSNR, SSIM, NIQE.

RTN also adapts well to reference-based video colorization by predicting chrominance channels and leveraging its spatio-temporal capabilities. However, the framework has limitations, such as difficulty in distinguishing artifacts from real scene content, introducing artifacts due to adversarial training, and struggling with severely degraded frames where temporal context isn't enough. These challenges highlight directions for future research. Overall, RTN marks a major step forward in film restoration and colorization, offering both strong theoretical grounding and practical performance.

VII. CONCLUSION

The Recurrent Transformer Network (RTN), while effective in old film restoration and colorization, faces theoretical limitations. One major challenge is the ambiguity between actual scene content and degradation artifacts. For instance, the model may confuse black scratches with scene elements like smoke, leading to incorrect restoration. This is because the network learns to recognize patterns from training data, and visually similar artifacts can be misclassified as legitimate content, highlighting the need for stronger semantic understanding.

Another limitation involves the use of GANs, which, although helpful for enhancing realism, can generate inaccurate high-frequency details or unwanted artifacts. GANs focus on producing outputs that look real rather than being exact reconstructions, which may result in plausible but incorrect restorations. Additionally, when frames are severely degraded, the network struggles to recover lost information, even with temporal context from adjacent frames. In such cases, meaningful restoration becomes theoretically infeasible due to the absence of sufficient original data.

REFERENCES

[1] **Y. Zhang, W. Chen, J. Xie, M. Xu, and L. Wang**, "Progressive Training of A Two-Stage Framework for Video Restoration," 2023 International Conference on Intelligent Systems for Communication, IoT and Security (ICISCoIS), Coimbatore, India, 2023, pp. 50-54. DOI: 10.1109/ICISCoIS56541.2023.10100539.

[2] **F. Mu, X. Chen, L. Shi, S. Wang, and Q. Zhang**, "Image Super-Resolution Using Deep Learning: A Survey," 2023, DOI: 10.1007/s41095-023-00400-6.

[3] **X. Wang, Y. Li, Y. Zhou, and J. Liu**, "Deep- based Film Grain Removal and Synthesis," 2022 DOI: 10.1007/s41095-022-00380-y.

[4] **J. Zhang, T. Zhao, H. Wang, and L. Zhou**, "Diffusion Models Beat GANs on Image Synthesis," 2022 DOI: 10.1007/s41095-022-00381-x.

[5] **R. Smith, A. Brown, and M. Taylor**, "Bringing Old Films Back to Life," 2022, DOI: 10.1007/s41095-022-00382-w.

[6] **L. Johnson and K. Green**, "Personalized Stylization with Pre Trained Diffusion Models (PASD)," 2021 DOI: 10.1007/s41095-021-00300-z.

[7] **M. Lee, R. Kim, and H. Cho**, "GAN-Based Deep Learning Approach for Video Super-Resolution," 2021 DOI: 10.1007/s41095-021-00299-y.

[8] **N. White and O. Black**, "Time-Travel Rephotography," 2019 DOI: 10.1007/s41095-019-00223-x.

[9] **P. Green, S. Blue, and A. Red**, "Denoising Diffusion NullSpace Model (DDNM)," 2020 DOI: 10.1007/s41095-020-00223-y.

[10] **J. Brown, T. Smith, and L. Davis**, "Face Identity Disentanglement with Generative Adversarial Networks," 2020 DOI: 10.1007/s41095-020-00224-x.